Appendix A

Documentation

Section 1. Documentation Guide

Appendix A of the *State Energy Data Report* describes how the estimates in the report were derived by the Combined State Energy Data System (CSEDS). The following five sections, one for each energy source, provide: descriptions of all the data series that are entered into CSEDS; the formulas applied in CSEDS for creating additional data series; and notes on special circumstances for any series.

Appendix B is an alphabetical listing of the variable names and formulas used in the system; Appendix C lists the conversion factors used in CSEDS to convert physical units into British thermal units and gives the sources for those factors; Appendix D provides the U.S. Department of Commerce, Bureau of the Census, resident population data used in per capita calculations; Appendix E presents metric and other physical conversion factors for information, although they are not currently used in CSEDS; Appendix F lists carbon dioxide emission factors for coal consumed by State for information, although they are not used in CSEDS; Appendix G is a summary of the changes made in CSEDS since the last report, which was released in December 1997; and Appendix H is a list of other Energy Information Administration reports containing State-level data.

There are 475 variables used in CSEDS to create the estimates in this report. All of the variables are identified by seven-letter names, such as MGTCPAL. In the following example, MGTCPAL is the identifying code for data on motor gasoline total consumption in physical units in Alabama

Characters:	**MG**	**TC**	**P**	**AL**
Positions:	1 and 2	3 and 4	5	6 and 7
Identity:	Type of Energy	Energy activity or consumption end-use sector	Type of data	Geographic

The type of energy categories in CSEDS, which are represented by the first two letters of the variable name, are:

AB = aviation gasoline blending components
AC = anthracite
AI = aluminum ingot
AR = asphalt and road oil
AS = asphalt
AV = aviation gasoline
BC = bituminous coal and lignite
BM = biomass
CC = coal coke
CG = corrugated and solid fiber boxes
CL = coal
CO = crude oil, including lease condensate
CT = catalytic cracking
DF = distillate fuel
DK = distillate fuel, including kerosene-type jet fuel

EL = electricity
EN = ethanol
ER = electricity generated from renewable energy
ES = electricity sales
EX = electricity generated from non-renewable energy
FF = fossil fuels
FN = petrochemical feedstocks, naphtha less than 401° F
FO = petrochemical feedstocks, other oils equal to or greater than 401° F
FS = petrochemical feedstocks, still gas
GE = geothermal energy
GO = geothermal, wind, photovoltaic, and solar thermal energy
HP = hydroelectric power from pumped storage
HV = conventional hydroelectric power
HY = hydroelectric power, all types
JF = jet fuel
JK = jet fuel, kerosene-type
JN = jet fuel, naphtha-type
KS = kerosene
LG = liquefied petroleum gases
LO = electrical system energy losses
LU = lubricants
MB = motor gasoline blending components
MG = motor gasoline
MS = miscellaneous petroleum products
NA = natural gasoline (including isopentane)
NG = natural gas
NU = nuclear electric power
OC = organic chemicals
PA = all petroleum products
PC = petroleum coke
PI = paints and allied products
PL = plant condensate
PO = other petroleum products
PP = pentanes plus
RD = road oil
RE = renewable energy
RF = residual fuel
SG = still gas
SN = special naphtha
SO = photovoltaic and solar thermal energy
TE = total energy
TN = total net energy
TP = resident population
UO = unfinished oils
US = unfractionated stream
WD = wood
WN = wind, photovoltaic, and solar thermal energy
WS = waste
WW = wood and waste
WX = waxes
WY = wind

The consumption end-use sectors, identified by characters three and four of each variable name, such as:

AC = transportation sector consumption
CC = commercial sector consumption
EU = electric utility sector consumption
IC = industrial sector consumption
RC = residential sector consumption
TC = total consumption of all sectors

Many other characters occur in the third and fourth positions of the variable names for the sales, deliveries, and distribution data series used in the intermediate calculations in CSEDS to derive the end-use consumption estimates. Examples of these codes are:

AG = sales for use in agriculture
BK = sales for use in vessel bunkering
IN = deliveries to the industrial sector
OD = distribution to other industrial users

Combining the first two components (the first four letters) produces variable names, such as:

MGAG = motor gasoline sold for use in agriculture
MGAC = motor gasoline consumed by the transportation sector
NGIN = natural gas delivered to the industrial sector
NGIC = natural gas consumed by the industrial sector

The fifth character of the variable names in CSEDS identifies the type of data by using one of the following letters:

B = data in British thermal units (Btu)
K = factor for converting data from physical units to Btu
M = data in alternative physical units
P = data in standardized physical units
S = share or ratio expressed as a fraction
V = value added in manufacture

Data entered into CSEDS are in physical units, represented by a "P" in the fifth character; for example, coal data are in thousand short tons, petroleum data are in thousand barrels, and natural gas data are in million cubic feet. In a few cases, data are obtained from the source documents in different units, such as thousand gallons instead of thousand barrels, and are represented by an "M" until converted in CSEDS to the unit that is consistent with other variables. Conversion factors, represented by a "K" in the fifth character, are applied to the physical unit data to convert the data to British thermal units, a common unit for all forms of energy. The derived data series in thousand British thermal units are represented by "B" in the fifth character. In a few cases, consumption estimates are derived by calculating shares of aggregated consumption data. The fractions used to calculate the consumption shares are identified by an "S" in the fifth character. The consumption estimates for some petroleum products are based on the value added in the manufacturing process by related industries in each State. The data series for those industry activities are in dollars, and the variable names contain "V" in the fifth character.

The last two characters of each variable name are for geographic identification. Geographic areas used in CSEDS are the 50 States and the District of Columbia (represented by the U.S. Postal Service State abbreviations) and the United States as a whole. Some estimates of electricity sales and losses are derived by using only the contiguous 48 States and the District of Columbia, and the variables used in those calculations are identified by "48" in the last two characters of the names. The geographic area codes used in CSEDS are shown in Table A1.

Table A1. Geographic Area Codes Used in the State Energy Data System

Code	State	Code	State
AK	Alaska	NC	North Carolina
AL	Alabama	ND	North Dakota
AR	Arkansas	NE	Nebraska
AZ	Arizona	NH	New Hampshire
CA	California	NJ	New Jersey
CO	Colorado	NM	New Mexico
CT	Connecticut	NV	Nevada
DC	District of Columbia	NY	New York
DE	Delaware	OH	Ohio
FL	Florida	OK	Oklahoma
GA	Georgia	OR	Oregon
HI	Hawaii	PA	Pennsylvania
IA	Iowa	RI	Rhode Island
ID	Idaho	SC	South Carolina
IL	Illinois	SD	South Dakota
IN	Indiana	TN	Tennessee
KS	Kansas	TX	Texas
KY	Kentucky	UT	Utah
LA	Louisiana	VA	Virginia
MA	Massachusetts	VT	Vermont
MD	Maryland	WA	Washington
ME	Maine	WI	Wisconsin
MI	Michigan	WV	West Virginia
MN	Minnesota	WY	Wyoming
MO	Missouri	US	United States
MS	Mississippi	48	The contiguous 48 States and the District of Columbia
MT	Montana		

Throughout this report, the term "State" includes the District of Columbia. Throughout this documentation, "ZZ" is used as a geographic identifier to represent the different State abbreviations that would be interchanged in that position of the variable name.